Critical Review Form
Diagnostic Test

Ovarian Torsion: A Fifteen-Year Review, *Ann Emerg Med* 2001; 38: 156-159

Objective: **"To describe the history, physical, and laboratory findings in women with ovarian torsion (OT)". (p. 159)**

Methods: **Retrospective chart review at two urban residency training hospitals in Denver CO. Women admitted between January 1984 and December 1999 were eligible for inclusion if they had surgically proven ovarian torsion (OT). Using a standardized review form, a single trained data abstractor, who was not blinded to the study purpose or patients' final diagnosis, obtained data elements from ED (resident and attending) physician notes, surgical or OB/GYN notes, and resident admission history and physical examination. Data elements that were recorded included pain descriptors, prior surgeries, fertility status, history of ovarian cysts, pain duration/radiation/associated symptoms. Another author also reviewed 10 randomly selected charts for interrater reliability (The $\kappa = 0.778$, although authors do not describe whether this is overall data elements or for some particular subset of the chart review). The absence of particular data elements were coded as negative (not present). (p. 157)**

Guide		Comments
I.	**Are the results valid?**	
A.	**Did clinicians face diagnostic uncertainty?**	Yes. Female patients presenting (mostly to the ED) with undifferentiated abdominal pain.

B.	**Was there a blind comparison with an independent gold standard applied similarly to the treatment group and to the control group?** **(Confirmation Bias)**	Yes. "Women with a diagnosis of OT who did not have torsion confirmed surgically were not included in this study". (p. 157) "Only surgically proven cases of OT were included in this study, and we may have missed patients with less "typical" presentations of OT who did not have surgery and therefore were not diagnosed with OT". (p. 157) In general, a less biased study design would have been prospective enrollment of consecutive females with abdominal pain with pre-established criteria for data collection before the actual clinical evaluation (so that the same methods are used to obtain the same elements of history & physical examination on each patient). Such a design would permit construction of 2x2 table's to calculate likelihood ratios, and a fuller understanding of the diagnostic accuracy for typical signs, symptoms, and imaging modalities. Unfortunately, such trials for rare conditions (six cases a year in these two hospitals) would be expensive and time-consuming to conduct. Therefore, none currently exist.
C.	**Did the results of the test being evaluated influence the decision to perform the gold standard?** **(Ascertainment Bias)**	Yes, undoubtedly the findings on many of the history and physical exam elements being assessed impacted surgeons' decisions to pursue laparoscopy to yield the criterion standard surgically confirmed OT.
II.	**What are the results?**	

<table>
<tr>
<td>A.</td>
<td>What likelihood ratios were associated with the range of possible test results?

<u>Sensitivity of History</u>
<table>
<tr><th>Characteristic</th><th># pts</th><th>Sens. (%)</th></tr>
<tr><td>Sudden onset</td><td>51</td><td>59</td></tr>
<tr><td>Prior pelvic surgery</td><td>35</td><td>40</td></tr>
<tr><td>Prior ovarian cyst</td><td>22</td><td>25</td></tr>
<tr><td>Hysterectomy</td><td>7</td><td>8</td></tr>
<tr><td>Prior pelvic disease</td><td>13</td><td>15</td></tr>
</table>

<u>Sensitivity of Signs</u>
<table>
<tr><th>Characteristic</th><th># pts</th><th>Sens. (%)</th></tr>
<tr><td>Nausea and/or vomiting</td><td>61</td><td>70</td></tr>
<tr><td>Sharp/stabbing pain</td><td>61</td><td>70</td></tr>
<tr><td>Crampy/colicky pain</td><td>38</td><td>44</td></tr>
<tr><td>Radiation (pain, flank, groin)</td><td>44</td><td>51</td></tr>
<tr><td>Lower quadrant pain</td><td>80</td><td>90</td></tr>
<tr><td>Prior pain episodes</td><td>37</td><td>43</td></tr>
<tr><td>Moderate to severe pain</td><td>71</td><td>82</td></tr>
</table>

<u>Sensitivity of Physical Exam</u>
<table>
<tr><th>Characteristic</th><th># pts</th><th>Sens. (%)</th></tr>
<tr><td>Mild tenderness abd exam</td><td>30</td><td>35</td></tr>
<tr><td>No tenderness pelvic exam</td><td>25</td><td>29</td></tr>
<tr><td>Palpable pelvic mass</td><td>41</td><td>47</td></tr>
</table>
</td>
<td>

Although 101 charts were identified by discharge coding as OT, only 87 met eligibility criteria (5 had no surgical OT, 2 lacked original encounter records, 2 left AMA, and 5 charts not found).
Among the 87 meeting eligibility criteria, age ranged 14-82 years and 75% presented to an ED first (10% presented to an OB/GYN clinic). Additionally, 17% were post-menopausal and 14% was pregnant.
Median time from pain onset was 1 day (mean 7.8 days) and 45% presented to the ED within 12 hours of pain onset.
Thirteen patients (16%) had WBC > 15,000.
93% (70/75) had abnormal ultrasound results (specific abnormalities not detailed) including 6/9 with abnormal Doppler flow.
OT was considered in the admission differential diagnosis in 47% and only 26 had surgery within 24 hours (mean time from presentation to surgery was 5.8 days).
89% had an ovary > 5cm.
The three most commonly associated ovarian pathologies were hemorrhagic cyst (29%), benign teratoma (22%) and serous cystadenoma (13%).

</td>
</tr>
<tr>
<td>III.</td>
<td>How can I apply the results to patient care?</td>
<td></td>
</tr>
<tr>
<td>A.</td>
<td>Will the reproducibility of the test result and its interpretation be satisfactory in my clinical setting?</td>
<td>Uncertain since the authors do not provide estimates of interrater reliability for findings from history or physical exam nor do they provide CI’s for estimates of sensitivity.</td>
</tr>
</table>

B.	**Are the results applicable to the patients in my practice?**	Yes, 75% of these patients presented first to the ED. The analysis would have been substantially more useful for EP's if stratified by patients presenting to ED first and sensitivity of diagnostic variables as assessed by EM.
C.	**Will the results change my management strategy?**	Yes, by recognizing that: a) History and physical exam are not sensitive for OT b) OT is a rare diagnosis c) Ultrasound and Doppler studies are not 100% sensitive for OT.
D.	**Will patients be better off as a result of the test?**	Unknown. No patient-centric outcomes (time to relief of pain, functional ovarian salvage) were reported. However, "the main significance of diagnosing OT may rest in the exclusion of other diagnoses". (p. 159)

Limitations

1) **Failure to report sufficient detail for chart review methods. Specifically (Worster 2004):**
   a) **How to assess data quality. What proportion of data was missing?**
   b) **Since more than one clinician's charts were being reviewed, how were conflicting data coded?**
   c) **What was the overall prevalence of OT during this time interval and what sample size would be needed to optimize assessment of history/physical exam diagnostic accuracy?**
   d) **How cases were identified (i.e. sampling method from hospital vs. ED discharge coding)?**

2) **Failure to report 95% CI around sensitivity point estimates.**

3) **Failure to stratify analysis by patient site of entry (ED, non-ED) or physician obtaining data (EM vs. non-EM).**

4) **Retrospective case series design without ability to compute specificity or LR's.**

5) **Failure to assess sensitivity for constellations of symptoms.**

6) **No explicit description of what surgical criteria were used to diagnose OT.**

Bottom Line

Lower quadrant pain (90% sensitive) of moderate to severe intensity (82% sensitive) may be useful in the diagnosis of OT, but all other signs, symptoms, and physical exam findings are not sufficiently sensitive to be useful in EM. The specificity of history/physical exam findings, as well as the reproducibility, remains undefined. Based upon these results, EP's should maintain OT in the differential diagnosis of women of all ages with abdominal pain (as long as they still have their ovaries) and maintain a low threshold for further imaging (US with Doppler, CT, or MRI) and GYN consult while recognizing that these modalities are also imperfect. Certainty may only reside in laparoscopy, although future trials should assess patient-important outcomes in addition to full (sensitivity and specificity, likelihood ratios) diagnostic accuracy.